SC-RoadDeepNet: A New Shape and Connectivity-preserving Road Extraction Deep Learning-based Network from Remote Sensing Data

Abolfazl Abdollahi[1], Biswajeet Pradhan[1,2,3,*], Abdullah Alamri[4]

***Abstract*—Existing automated road extraction approaches concentrate on regional accuracy rather than road shape and connectivity quality. Most of these techniques produce discontinuous outputs caused by obstacles, such as shadows, buildings, and vehicles. This study proposes a shape and connectivity-preserving road identification deep learning-based architecture called SC-RoadDeepNet to overcome the discontinuous results and the quality of road shape and connectivity. The proposed model comprises a state-of-the-art deep learning-based network, namely, the recurrent residual convolutional neural network, boundary learning (BL), and a new measure based on the intersection of segmentation masks and their (morphological) skeleton called connectivity-preserving centerline Dice (CP_clDice). The recurrent residual convolutional layers accumulate low-level features for segmentation tasks, thus allowing for better feature representation. Such representation enables us to construct a UNet network with the same number of network parameters but improved segmentation effectiveness. BL also aids the model in improving the road's boundaries by penalizing boundary misclassification and fine-tuning the road form. Furthermore, the CP_clDice method aids the model in maintaining road connectivity and obtaining accurate segmentations. We demonstrate that CP_clDice ensures connection preservation for binary segmentation, thereby allowing for efficient road network extraction at the end. The proposed model improves F1 score accuracy to 5.49%, 4.03%, 3.42%, and 2.27% compared with other comparative models, such as LinkNet, ResUNet, UNet, and VNet, respectively. Furthermore, qualitative and quantitative assessments demonstrate that the proposed SC-RoadDeepNet can improve road extraction by tackling shadow and occlusion-related interruptions. These assessments can also produce high-resolution results, particularly in the area of road network completeness.**

***Index Terms*— Deep learning; remote sensing; road extraction; road shape and connectivity preservation**

[1]A. Abdollahi and B. Pradhan are with the Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), School of Civil and Environmental Engineering, Faculty of Engineering and IT, University of Technology Sydney, 2007 Sydney, NSW, Australia.

[2]B. Pradhan is also with the Department of Energy and Mineral Resources Engineering, Sejong University, Choongmu-gwan, 209 Neungdong-ro, Gwangjin-gu, Seoul, 05006, Republic of Korea.

[3]B. Pradhan is also with the Earth Observation Centre, Institute of Climate Change, Universiti Kebangsaan Malaysia, 43600 UKM, Bangi, Selangor, Malaysia

[4]A. Alamri is with the Department of Geology and Geophysics, College of Science, King Saud University, P.O. Box 2455, Riyadh 11451

(Correspondence: biswajeet24@gmail.com or Biswajeet.Pradhan@uts.edu.au).

I. INTRODUCTION

Very-high-resolution (VHR) images have become a crucial geospatial data source because of their extensive coverage and high accuracy [1]. The road network information derived from these imageries is useful in various applications containing transportation systems development, cartography, urban planning, and navigation [2]. Road networks form the majority of modern transportation infrastructure because they are significant man-made ground objects. Roads also provide essential data in geographic information systems; thus, their timely updates can impact numerous applications (e.g., emergency response and route analysis) that rely on these datasets [3].

The most common method of extracting roads has been through manual visual interpretation, which takes a long time and costs a lot of money. Moreover, the obtained outcomes may differ because of the interpreter's discrepancies. The technology of automatic road extraction has been a popular topic in this field because it can increase the effectiveness of road extraction [4]. However, high-resolution imagery can reveal the vehicles on the road and the shadows of buildings or trees on the roadside. Furthermore, the road segments are irregular, and the roads structures are complex [5]. The abovementioned challenges make extracting road networks from high-resolution data more difficult [6].

Some scholars have used traditional methods or machine learning algorithms to overcome these difficulties, as evidenced by substantial studies in the literature. For example, a semi-automatic approach based on mean shift was presented by [7] to extract roads. The method separates the boundary between non-roads and roads by extracting the initial point from road seed points and a threshold. Furthermore, Unsalan and Sirmacek [8] applied graph theory and probability for road network extraction. In addition, Bakhtiari et al. [9] implemented a semi-automatic method based on edge detection, support vector machine, and morphological operations to extract roads from VHR imagery. Compared with the methods mentioned above, machine learning approaches are usually more accurate. For instance, Alshehhi and Marpu [10] suggested a hierarchical graph-based image segmentation strategy for road extraction. Das et al. [11] extracted road networks from high-resolution multispectral imagery based on designing a multistage framework to exploit two salient road features. Song and Civco [12] used SVM and shape index features to extract road sections. Although these methods may work well in some simple circumstances, their effectiveness is dependent on several threshold criteria that must be specified elaborately. Given that threshold settings fluctuate between imagery, conventional approaches can only perform with a limited set of data and cannot be tested in complex environments [13].

The deep learning technique, which is characterized by convolutional neural networks (CNNs), has attained a milestone in the computer vision field, owing to the exponential development of accessible data and computational capacity [14-19]. Researchers have preferred to use CNN-based algorithms to extract roads from remote sensing data in recent years because road extraction can be regarded as a binary segmentation issue. Mnih and Hinton [20] proposed a CNN method to extract roads from aerial imagery in early 2013. Furthermore, Rezaee and Zhang [21] developed a patch-based CNN model for extracting roads from images with 0.15 m spatial resolution. In another study, Wang et al. [5] used a patch-based CNN and finite state machine (FSM) model to recognize road patterns and track roads. These patch-based techniques use a sliding window technique, which limits their speed and efficiency. The road detection problem has made significant progress [22] with the advent of a significant number of outstanding semantic segmentation structures based on encoder-decoder frameworks, including DeepLab [23], UNet [24], and SegNet [25], or a fully convolutional network (FCN) [26]. Li et al. [27] detected a road from unmanned aerial vehicle imagery (UAV) using an improved D-LinkNet model. Meanwhile, Zhang [28] built a deep residual UNet (ResUNet) for road detection, which incorporates UNet with residual units in its architecture. To provide a wide receptive field, Zhang and Wang [29] presented a network with atrous convolution, which functions well in building and road extraction. Furthermore, Zhong et al. [30] developed an FCN model for road extraction that integrates the deep final-score layer with the shallow fine-grained pooling layer output.

Several works have updated the loss function to produce better road extraction outcomes and improve the network structure. For example, to increase the quality of road extraction, He et al. [31] used structural similarity as a loss function. Furthermore, to reduce class imbalance and improve the road extraction results, Abdollahi et al. [32] performed a VNet network with a novel combined loss function named the cross-entropy-dice-loss (CEDL) function. Moreover, Mosinska et al. [33] applied a pixel-wise loss function to preserve the topological characteristics of roads structures.

All the approaches listed above can reliably segment roads in remote sensing imagery; nevertheless, they fail to detect roads obscured by buildings, shadows, trees, or other non-road features [13]. Given the complex characteristics of covered roads, typical FCNs-based approaches cannot detect them accurately. Furthermore, given that these techniques are mainly encoder-decoder architectures, the boundary precision of the road extraction findings will diminish during the downsampling phase [34]. The number of feature maps in the encoder rises as the model goes deeper, whereas the spatial resolution declines [34]. The spatial resolution of feature maps is gradually recovered in the decoder arm through the up-sampling layer. However, edge information is lost through the process. Given that roads are man-made objects with distinct borders, concentrating on boundary and topology precision increases road network quality. Conventional FCN-based approaches convey context information through convolutional and down-sampling operations in the local receptive fields. Thus, they experience difficulties when detecting roads obscured by trees or buildings. The context information modeling mechanisms of traditional FCNs cannot build topological links between road segments split by obstacles, thus resulting in fragmented and discontinuous results for road extraction. Therefore, to address the challenges in shape accuracy and connectivity, a shape and connectivity-preserving road detection deep learning-based architecture (SC-RoadDeepNet) is suggested in this study.

In the proposed model, we implement a new deep learning-based network called the recurrent residual CNN model (RRCNN), which is based on the UNet network. The presented network uses recurrent residual convolutional layers (RRCLs), UNet, and residual networks. For segmentation tasks, RRCLs accumulate important features and thus enable better feature representation. They allow us to build a UNet network with similar network parameters but better segmentation performance. We also use road boundaries to make road semantic features more proper for the actual road form, solve irregular semantic features, and enhance the boundary of road semantic polygons. We leverage each road's binary edge-map to penalize boundary misclassification and fine-tune the road shape.

Furthermore, we offer a connectivity-preserving centerline Dice (CP_clDice), a new measure based on the intersection of segmentation masks and their (morphological) skeleton, to preserve road connectivity and obtain accurate segmentations. Our measure states the network's connectivity rather than evenly weighting each pixel given its morphological skeleton-based formulation. We show that CP_clDice ensures connectivity conservation for binary segmentation, thus allowing for proper road network extraction. We present experimental results on a challenging road dataset that includes original references and Google Earth images with a spatial resolution of 0.21 m per pixel, encompassing 21 urban regions of approximately 8 km^2 with complex backgrounds.

The rest of this paper is laid out as follows. An overview of the suggested method is introduced in Section II. Then, the comprehensive information about our Google Earth road dataset and experimental settings is described in Section III. The experimental results and ablation analyses are shown in Sections IV and V, respectively. Section VI presents the conclusion and main findings obtained in this study.

II. METHODOLOGY

This work suggests a new shape and connectivity-preserving road detection deep learning-based architecture (SC-RoadDeepNet) from Google Earth imagery. The proposed technique consists of a deep learning model named RRCNN based on the original UNet network with better performance, the binary edge-map of each road, and a new connectivity-aware similarity measure based on intersecting skeletons with masks (CP_clDice) to preserve road connectivity. In the following, the architecture of the RRCNN network and CP_clDice measure are explained.

A. The Architecture of RRCNN

We propose RRCNN (Figure 1), a new model for segmentation tasks that is inspired by UNet [24] (Figure 2), RCNN [35], and the deep residual model [36]. The original UNet model consists of two main parts: convolutional encoding and decoding units. In both the encoder and decoder parts of the model, the fundamental convolutional layers are applied, followed by ReLU activation. In the encoding part, 2×2 max-pooling layers are applied for down sampling [24]. The convolutional transpose layers are used to up-sample the feature maps during the decoding step. In the UNet network, cropping and copying method is used to crop and copy feature maps from the encoder part to the decoder part [24]. Therefore, the benefits of all three established deep learning approaches are combined in the proposed approach. Assuming a pixel in an input sample on the k^{th} feature map in the recurrent convolutional layers (RCL) that is located at (i,j) and input sample x_l in the layer l^{th} of the RCNN block, the network's output $o_{ijk}^{l}(t)$ at the t time step can be expressed as follows:

$$O_{ijk}^{l}(t)=(w_k^f)^T \times x_l^{f(i,j)}(t)+(w_k^r)^T \times x_l^{r(i,j)}(t-1)+b_k, \quad (1)$$

where b_k is the bias, w_k^r is the weight of the k^{th} RCL's feature map, w_k^f is the standard convolutional layer's weight, $x_l^{r(i,j)}(t-1)$ is the input for the l^{th} RCL, and $x_l^{f(i,j)}(t)$ is the input for the standard convolutional layers. The RCL's outputs are passed through the rectified linear unit (ReLU) activation function f, which is denoted as follows:

$$F(x_l w_l)=f(O_{ijk}^{l}(t))=\max(0,O_{ijk}^{l}(t)), \quad (2)$$

where $F(x_l w_l)$ denotes that the outputs of the l^{th} RCNN layer are used in the encoding and decoding arms of the network for down-sampling and up-sampling layers, respectively. For the RRCNN model, the last output that is passed through residual units can be expressed as follows:

$$x_{l+1}=x_l+F(x_l w_l), \quad (3)$$

where, in the RRCNN's encoding and decoding arms, x_{l+1} is used as the input for immediate subsequent down or up-sampling layers, and the RRCNN-input block's samples are represented by x_l.

The suggested RRCNN model is the building block of the stacked recurrent residual convolutional units depicted in Figure 3(c). This study investigated convolutional and recurrent convolutional units in various variants for three distinct architectures, as shown in Figures 3(a)–3(c). The first is the primary UNet architecture [24] with encoder-decoder arms and a crop and copy method (skip connection). This model's fundamental convolutional unit is depicted in Figure 3(a). The second is ResUNet [37], which is the original UNet model with forwarding convolutional and residual connection units, as illustrated in Figure 3(b).

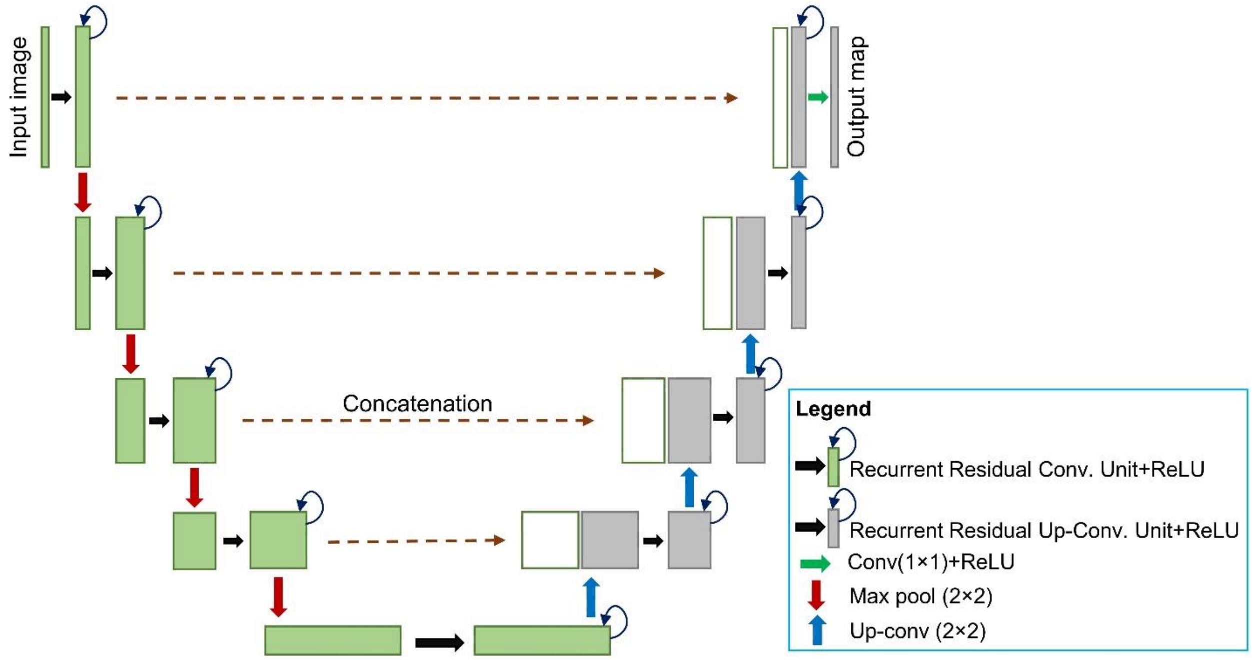


Fig. 1. Architecture of the proposed RRCNN model, including encoder-decoder units based on recurrent RRCL and UNet networks

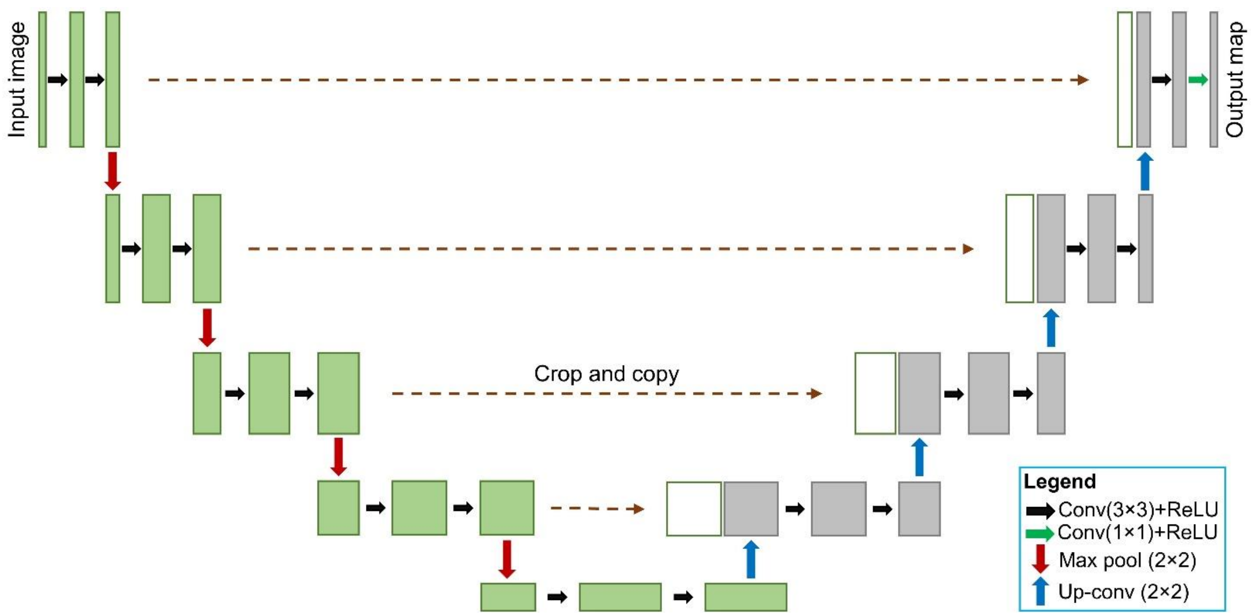


Fig. 2. Architecture of the original UNet model, including convolutional encoder-decoder units

The final architecture is the proposed RRCNN, including the primary UNet with RCL and residual connections, as depicted in Figure 3(c). When compared with UNet, the proposed architecture offers various advantages. One of these advantages is network productivity, which is measured in relation to the number of network parameters. Compared with UNet and ResUNet, the suggested RRCNN model is built to have similar parameters while performing efficiently on feature extraction. Recurrent or residual units do not increase the number of network parameters. However, they have a considerable effect on the training/testing results. Furthermore, the RCL units of the proposed model provide an efficient feature accumulation mechanism. Concerning distinct time-steps, feature accumulation guarantees more reliable and robust feature representation. As a result, it aids in the extraction of low-level features that are critical for feature extraction. This, we eliminate the cropping and copying method from the primary UNet network and replace it with concatenation operation, which leads to a considerably more elegant design with improved efficiency.

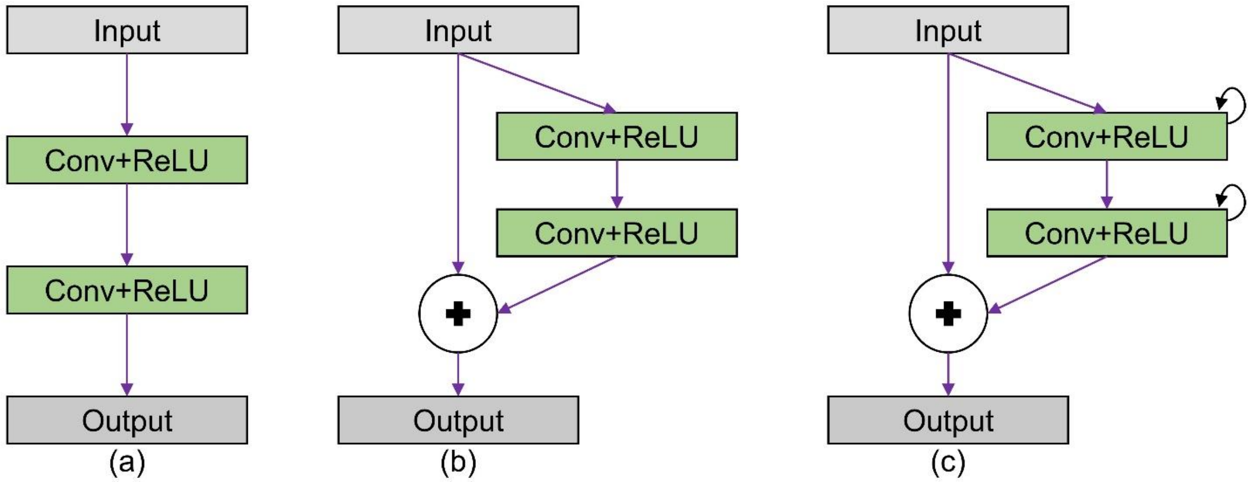


Fig. 3. Convolution and recurrent convolution units in various variants: (a) forward convolution units, (b) residual convolution units, and (c) recurrent residual convolution units.

B. Emphasizing Connectivity Using CP_clDice

Figure 4 depicts a schematic overview of our suggested CP_clDice technique. On the basis of intersecting skeletons with masks, we present a new connectivity-preserving measure for evaluating road structure segmentation. The ground truth (M_G) and detected segmentation (M_D) masks are two binary masks that we consider. From M_G and M_D, skeletons S_G and S_D are first extracted, respectively. $S_D = \{g_i\}_{i=1}^{N}$ is the detected skeleton of a detected mask M_D, while $S_G = \{h_i\}_{i=1}^{N}$ is the true skeleton of a true mask M_G, where h_i and g_i are the skeleton points of S_G and S_D, respectively. Then, we calculate the proportion of S_G that exists within M_D, which we call connectivity sensitivity or $C_{sens}(S_G, M_D)$, and vice-a-versa. We compute connectivity precision or $C_{prec}(S_D, M_G)$ as follows:

$$C_{sens}(S_G, M_D) = \frac{|S_G \cap M_D|}{|S_G|}; C_{prec}(S_D, M_G) = \frac{|S_D \cap M_G|}{|S_D|}, \quad (4)$$

Or $C_{sens} = \sum_{i=1}^{N} \frac{h_i M_D(h_i)}{\sum_{j=1}^{N} h_i}; C_{prec} = \sum_{i=1}^{N} \frac{g_i M_G(g_i)}{\sum_{j=1}^{N} g_i}$.

The metric the measure, $C_{sens}(S_G, M_D)$, is prone to false negatives in prediction, whereas $C_{prec}(S_D, M_G)$ is prone to false positives, thus clarifying why we refer to $C_{sens}(S_G, M_D)$ as the sensitivity of the connectivity and $C_{prec}(S_D, M_G)$ as its precision. We calculate CP_clDice as the harmonic mean of both measures because we want to maximize sensitivity and precision:

$$CP\_clDice(M_D, M_G) = 2 \times \frac{C_{prec}(S_D, M_G) \times C_{sens}(S_G, M_D)}{C_{prec}(S_D, M_G) + C_{sens}(S_G, M_D)}. \quad (5)$$

C. Soft-skeletonization with soft CP_clDice

The following section demonstrates how we use the CP_clDice formulation to train a connectivity-preserving network using our theory effectively. Our strategy relies on correct skeletons extraction. A variety of ways have been presented for this task. However, most of them are not entirely distinguishable and thus unsuitable for use in a loss function. The repeated morphological thinning [38] or Euclidean distance transform [39] are two popular methods. A series of erosions and dilation operations are used in morphological thinning. The Euclidean distance transform remains a discrete operation, thus prohibiting it from being used in a loss function for neural network training. As a grayscale alternative to morphological erosion and dilation, min and max filters are often used. As a result, we suggest soft-skeletonization, in which iterative min-max pooling is used as a surrogate for morphological dilation and erosion. Figures 5 and 6 illustrate the sequential steps of our skeletonization intuitively. Initial iterations (Figure 5) skeletonize and maintain structures with a small radius until later iterations skeletonize and maintain thicker structures, thus allowing for the creation of a parameter-free, morphologically focused soft skeleton. The iterative processes involved in its computation are described in Algorithm 1 (soft-skeletonization) shown in Figure 6. The iterations are represented by the hyper-parameter, which must be equal to or greater than the maximum witnessed radius.

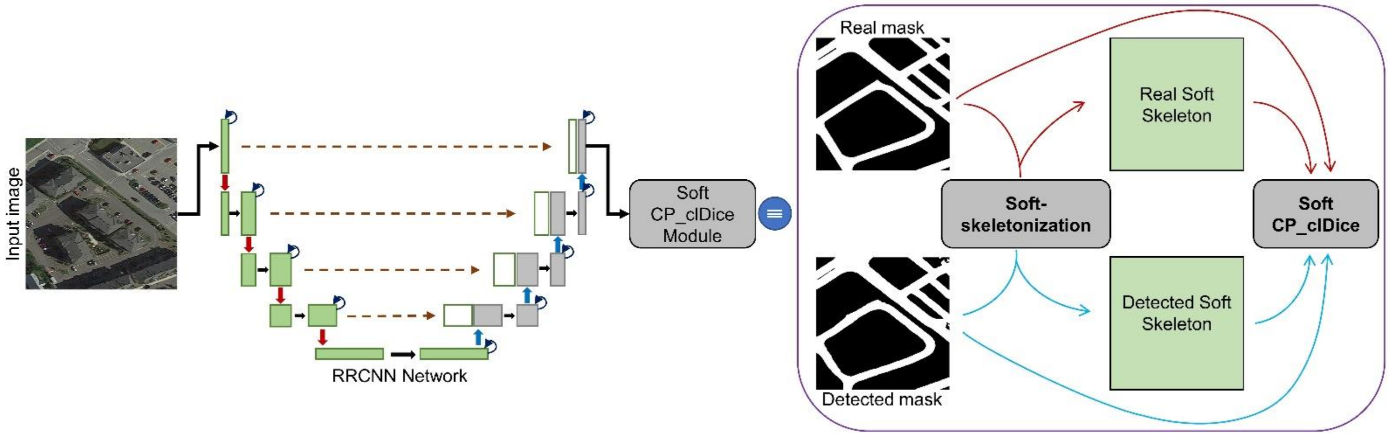


Fig. 4. An overview of our suggested CP_clDice technique. The CP_clDice method can be implemented in any generic segmentation model. We apply the RRCNN network in this work. Pooling functions from any common deep learning toolbox can be used to build soft-skeletonization.

This parameter varies depending on the dataset. For example, in our experiments, $k = 5...20$, which corresponds to the pixel radius of the largest witnessed road structures. A low k results in incomplete skeletonization. Increasing the value of k does not decrease the performance but lengthens the computation time. Given the previously stated soft-skeletonization, we can used CP_clDice as an optimizable, real-valued, and fully differentiable measure. The implementation is described in Algorithm 2 (Figure 6) and is known as the soft CP_clDice. The amount of linked loops determines the homotopy type for a single connected foreground component without knots. As a result, no pairwise linked loops are detected, and reference pixels are not homotopy-equal. The deformation retracted skeleton of the solid foreground must be added or removed to include or omit these extra loops. Thus, the addition of new pixels that have been appropriately detected is needed. Unlike other losses, such as cross-entropy and Dice, CP_clDice only analyzes the deformation-retracted graphs of the solid foreground structure. As a result, we assert that CP_clDice needs the minimum number of new properly detected pixels to ensure homotopy equality. Cross-entropy or Dice can only ensure homotopy equivalence in these lines provided that each pixel is properly segmented. CP_clDice can ensure the equivalence of homotopy for a wider combination of pixels, which is an intuitively appealing trait because it renders CP_clDice powerful against noisy segmentation labels.

D. Cost Function

We integrate our suggested soft CP_clDice with soft-Dice (a function to calculate dice loss) in the following manner to preserve connectivity while obtaining correct segmentations (our objective) rather than the learning skeleton:

$$L_c = (1-\alpha)(1-softDice) + \alpha(1-softCPclDice), \quad (6)$$

where $softDice = \dfrac{2\sum_i^N p_i o_i}{\sum_i^N p_i^2 + \sum_i^N o_i^2}$,

where N denotes the total pixels, $p_i \in M_D$ is the detected binary pixels, and $o_i \in M_G$ is the ground truth pixels.

This study aims to learn a connectivity-preserving segmentation, not learning the centerline. As a result, we limited α options (weight for the CP_clDice element) in our experiments to $[0.1, 0.5]$ to achieve high-quality results. Furthermore, we use the binary edge-map of each road to penalize boundary misclassification, solve irregular road forms, and enhance the shape of semantic roads. In fact, reliable annotated road edges are integrated into semantic polygons to improve the semantic polygon's border, repair discontinuous areas, assure the road's continuity and integrity, and obtain more precise boundary positioning. We test our CP_clDice and binary edge-map information on a new state-of-the-art deep learning model (RRCNN). We propose a new method named SC-RoadDeepNet, a shape and connectivity-preserving method, to show the effectiveness of the model in preserving connectivity while obtaining accurate segmentation.

III. EXPERIMENTS AND EVALUATION

We outlined the experimental dataset in-depth in this section. Then, we introduced the experimental setup in the suggested technique. Finally, we presented the evaluation measures used for assessing the accuracy of the proposed method.

A. Dataset

This part describes the dataset used to train and assess SC-RoadDeepNet, including Google Earth imagery [42], with a spatial resolution of 0.21 m per pixel covering approximately 8 km^2. The dataset was more comprehensive and difficult to work with because of the numerous obstacles and shadows generated by avenue trees and cars along the roads. A total of 696 images were included in the dataset, which was divided into a training set and a testing set of 651 images and 45 images. Every original image had a size of 512×512 pixels. Figure 7 shows various samples of the primary and corresponding ground truth imagery with different backgrounds in the dataset.

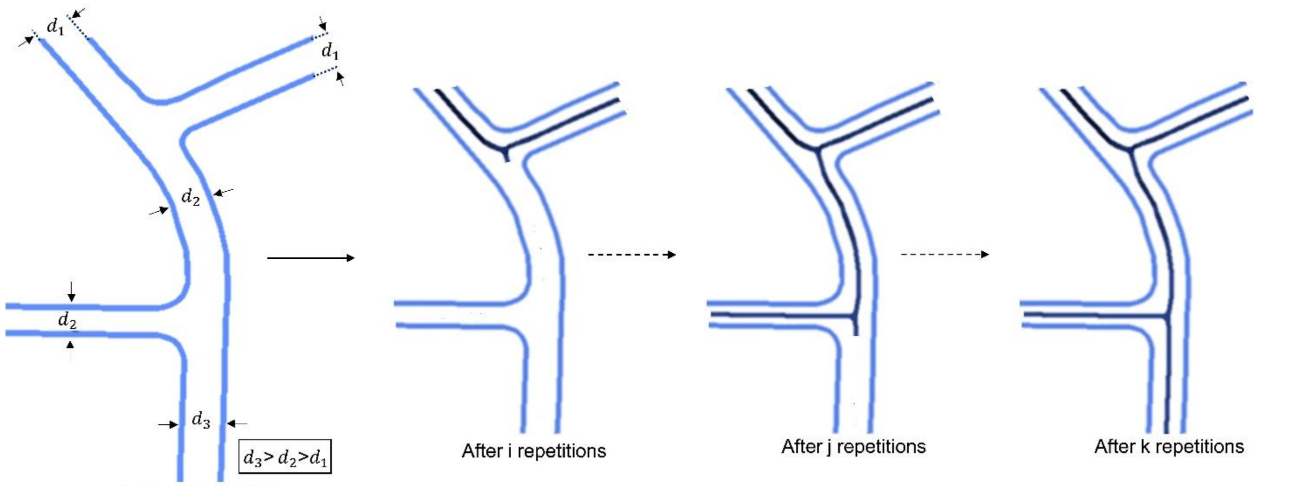


Fig. 5. Sequential bagging of skeleton pixels (dark blue) by iterative skeletonization leads to complete skeletonization based on the initial road structure (blue), where $k > j > i$ signifies iterations and d diameter.

Algorithm 1: soft-skeletonization

Input: M, k

$M' \leftarrow \max pooling(\min pooling(M))$

$Skel \leftarrow relu(M - M')$

for $m \leftarrow 0$ **to** k **do**

$M \leftarrow \min pooling(M)$

$M' \leftarrow \max pooling(\min pooling(M))$

$Skel \leftarrow Skel + (1 - Skel) \circ relu(M - M')$

end

Output: $Skel$

Algorithm 2: soft CP_clDice

Input: M_D, M_G

$S_D \leftarrow \textit{soft-skeletonization}\ (M_D)$

$S_G \leftarrow \textit{soft-skeletonization}\ (M_G)$

$$C_{prec}(S_D, M_G) \leftarrow \frac{|S_D \cap M_G|}{|S_D|}$$

$$C_{sens}(S_G, M_D) \leftarrow \frac{|S_G \cap M_D|}{|S_G|}$$

$CP\_clDice \leftarrow$

$$2\times\frac{C_{prec}(S_D, M_G)\times C_{sens}(S_G, M_D)}{C_{prec}(S_D, M_G)+C_{sens}(S_G, M_D)}$$

Output: *CP_clDice*

Fig. 6. The suggested soft-skeleton is calculated using Algorithm 1, where k is the number of iterations for skeletonization and M is the mask to be soft-skeletonized. The soft CP_clDice loss is calculated using Algorithm 2, where M_G is the ground truth mask and M_D is the segmentation mask. $\circ$ denotes the Hadamard product.

B. Experiment Settings

Given that the size of our road dataset was still small, which might lead to an over-fitting issue, some data augmentation techniques were utilized to increase the dataset size. We used data augmentation tactics, such as rotating (90, 180, and 270 degrees) the images and flipping (vertical and horizontal) them to enhance the dataset's capacity. The proposed network was trained on a GPU Nvidia Quadro RTX 6000 under Keras framework and with Tensorflow backend with batch size 1 for 100 epochs across the datasets. This study also used an adaptive moment estimation (Adam) optimizer with a $1e-3$ learning rate and decay of 0.9 to optimize the loss function and learn model parameters. The Sigmoid activation was also applied to sort the outcomes. The final layer provided outputs in the continuous value from 0 to 1, as it was activated by the Sigmoid function. As a result, we used a 0.5 threshold to attain the final segmentation map of the input images.

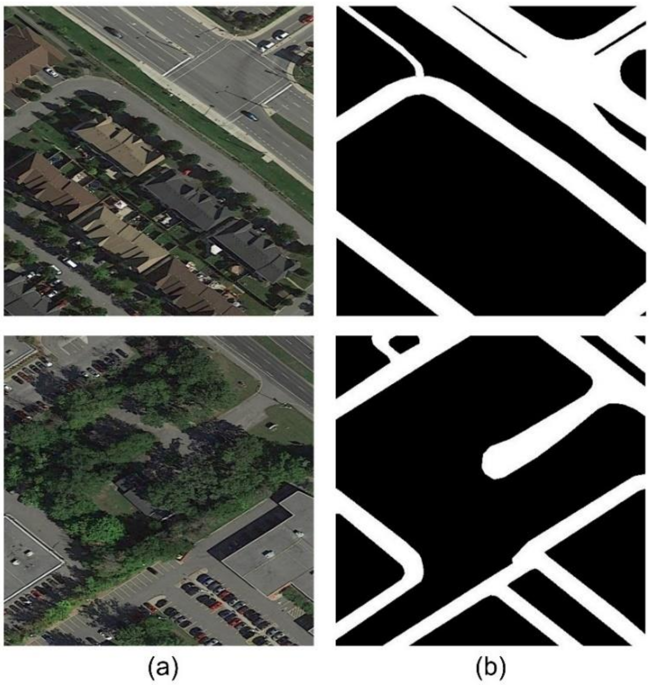


Fig. 7. Examples of (a) RGB Google Earth imagery and (b) their reference maps.

C. Evaluation Metrics

In this work, Precision, Recall, F1 score, Matthew Correlation Coefficient (MCC), Overall Accuracy (OA), and Intersection over Union (IoU) were used as metrics to analyze the suggested method's quantitative performance in road network extraction [40]. Precision and Recall came up with the F1 score. This score, which can be calculated as follows (7), is a powerful assessment metric for the harmonic mean of Precision and Recall.

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (7)$$ where $Precision = \frac{TP}{TP+FP}$ and $Recall = \frac{Tp}{TP+FN}$,

where the proportion of matched pixels in the extraction outcomes is measured by Precision and the percentage of matched pixels in the reference is measured by Recall. False negative, false positive, true positive, and true negative are represented by FN, FP, TP, and TN, respectively. The proportion of the overlapping predicted and reference areas to the whole area was measured by IoU (8), which is expressed as follows:

$$IoU = \frac{TP}{TP+FP+FN}. \quad (8)$$

MCC stands for the correlation coefficient between predicted and detected binary categorization, which is expressed as:

$$MCC = \frac{TP.TN - FP.FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}. \quad (9)$$

OA is also a simple summary assessment of a case's likelihood of being correctly classified, which is calculated as:

$$OA = \frac{TP+TN}{N}. \quad (10)$$

IV. EXPERIMENTAL RESULTS

This study was compared with some state-of-the-art techniques, including deep learning approaches, such as LinkNet [41], DeeplabV3+ [23], ResUNet [37], UNet [24], and VNet [32], to examine the applicability of the presented SC-RoadDeepNet method for road segmentation from Google Earth imagery. We tested the proposed RRCNN model by integrating the edge map to the semantic segmentation to determine how boundary learning (BL) fine-tunes the road shape via penalizing boundary misclassification. We called this network RRCNN-boundary-learning or RRCNN+BL. Furthermore, we compared the proposed SC-RoadDeepNet with different values of α, such as $\alpha = 0.1$, $\alpha = 0.3$, $\alpha = 0.5$, $\alpha = 0.7$, and $\alpha = 0.9$, to show the effect of the alpha parameter on the road connectivity and segmentation results. All of the mentioned methods were tried using the same collection of imagery to make the assessments fair and objective.

Table 1 demonstrates the quantitative findings obtained by the methods. The accuracy of the methods was calculated using IoU and F1 scores. LinkNet, DeeplabV3+, ResUNet, and UNet achieved the lowest IoU values with 81.52%, 82.53%, 84%, and 85.01%, respectively, when we compared the outcomes of different approaches (Table 1). VNet could improve the results to 86.99% compared with the mentioned four methods. By adding BL to the proposed RRCNN method (RRCNN+BL) and the proposed loss function to the model without BL (RRCNN+CP_clDIce), the accuracy of the IoU was also increased to 89.02% and 89.75%, respectively. These methods were the third-best and second-best methods in all approaches, thus proving the influence of edge-map and CP_clDice on improving road shape and segmentation results. In contrast, by including BL and connectivity-preserving CP_clDice techniques to the proposed SC-RoadDeepNet, IoU values reached 90.04%, 90.43%, 91.05%, 90.34%, and 89.85% for $\alpha = 0.1$, $\alpha = 0.3$, $\alpha = 0.5$, $\alpha = 0.7$ and $\alpha = 0.9$, respectively. We found that including CP_clDice in any value ($\alpha > 0$) resulted in the improvement of road connectivity and segmentation. Figures 8 and 9 also depict the qualitative results obtained using state-of-the-art techniques.

According to the findings, all extraction methods could reduce the impact of occlusions to some extent. However, LinkNet, DeeplabV3+, UNet, ResUNet, and VNet approaches were sensitive to noise and introduced some FPs in some parts, such as the shadows, buildings, and trees, and could not extract roads accurately. Benefiting from BL and CP_clDice, the proposed RRCNN+BL and RRCNN+CP_clDice methods could reduce boundary misclassification and achieve relatively satisfactory results. Furthermore, the proposed SC-RoadDeepNet, which took advantage of BL and CP_clDice techniques, could obtain fewer FPs (shown in blue) and FNs (shown in red), reduce road discontinuity, and produce high-resolution road segmentation maps compared with the other approaches. The presented SC-RoadDeepNet model with $\alpha = 0.5$ improved the results of IoU to 2.03% and 1.3% compared with the RRCNN+BL (third best) and RRCNN+CP_clDice (second best) models, respectively. They all showed that combining the suggested BL and CP_clDice techniques in the shape and connectivity-aware SC-RoadDeepNet model resulted in superior performance than other current approaches.

TABLE 1. QUANTITATIVE EXPERIMENTAL OUTCOMES YIELDED BY THE COMPARATIVE APPROACHES FOR THE GOOGLE EARTH ROAD DATASET.

		Image 1	Image 2	Image 3	Image 4	Image 5	Image 6	Average
LinkNet	F1 score	0.8821	0.9183	0.9149	0.8830	0.8970	0.8930	0.8981
	IoU	0.7890	0.8488	0.8430	0.7905	0.8132	0.8067	0.8152
	MCC	0.7903	0.8474	0.8615	0.8225	0.8341	0.8214	0.8295
	OA	0.8869	0.9231	0.9334	0.9154	0.9219	0.9137	0.9157
ResUNet	F1 score	0.8851	0.9302	0.9404	0.8870	0.9177	0.9157	0.9127
	IoU	0.7938	0.8694	0.8874	0.7969	0.8478	0.8445	0.8400
	MCC	0.7941	0.8662	0.9054	0.8288	0.8661	0.8564	0.8528
	OA	0.8923	0.9331	0.9556	0.9181	0.9364	0.9304	0.9277
UNet	F1 score	0.8901	0.9354	0.9313	0.9064	0.9284	0.9210	0.9188
	IoU	0.8019	0.8785	0.8714	0.8289	0.8663	0.8536	0.8501
	MCC	0.8051	0.8743	0.8906	0.8584	0.8840	0.8639	0.8627
	OA	0.8953	0.9372	0.9487	0.9333	0.9452	0.9328	0.9321
DeeplabV3+	F1 score	0.8626	0.9159	0.9254	0.8901	0.9226	0.9067	0.9039
	IoU	0.7584	0.8448	0.8612	0.8020	0.8563	0.8293	0.8253
	MCC	0.7516	0.8510	0.8791	0.8336	0.8668	0.8479	0.8383
	OA	0.8738	0.9231	0.9426	0.9206	0.9347	0.9274	0.9204
VNet	F1 score	0.9315	0.9390	0.9418	0.9108	0.9312	0.9277	0.9303
	IoU	0.8718	0.8850	0.8899	0.8361	0.8713	0.8650	0.8699
	MCC	0.8784	0.8797	0.9063	0.8647	0.8880	0.8758	0.8822
	OA	0.9382	0.9386	0.9559	0.9370	0.9463	0.9393	0.9426
RRCNN+BL	F1 score	0.9344	0.9517	0.9584	0.9209	0.9455	0.9397	0.9418
	IoU	0.8768	0.9078	0.9202	0.8534	0.8965	0.8862	0.8902
	MCC	0.8833	0.9052	0.9337	0.8811	0.9113	0.8971	0.9020
	OA	0.9414	0.9052	0.9689	0.9435	0.9576	0.9501	0.9445
RRCNN+CP_clDice	F1 score	0.9362	0.9669	0.9628	0.9213	0.9456	0.9418	0.9458
	IoU	0.8800	0.9359	0.9282	0.8541	0.8967	0.8900	0.8975
	MCC	0.8865	0.9352	0.9405	0.8810	0.9117	0.9002	0.9092
	OA	0.9423	0.9673	0.9721	0.9444	0.9578	0.9513	0.9559
SC-RoadDeepNet (α=0.1)	F1 score	0.9399	0.9719	0.9601	0.9247	0.9440	0.9437	0.9474
	IoU	0.8866	0.9453	0.9232	0.8599	0.8939	0.8934	0.9004
	MCC	0.8935	0.9450	0.9361	0.8865	0.9085	0.9034	0.9122
	OA	0.9462	0.9723	0.9700	0.9466	0.9557	0.9527	0.9573
SC-RoadDeepNet (α=0.3)	F1 score	0.9411	0.9726	0.9610	0.9301	0.9459	0.9467	0.9496
	IoU	0.8888	0.9466	0.9248	0.8693	0.8973	0.8988	0.9043
	MCC	0.8956	0.9479	0.9375	0.8945	0.9119	0.9090	0.9161

	OA	0.9472	0.9738	0.9610	0.9509	0.9575	0.9558	0.9577
SC-RoadDeepNet (α=0.5)	F1 score	0.9435	0.9775	0.9677	0.9331	0.9466	0.9493	0.9530
	IoU	0.8929	0.9560	0.9374	0.8746	0.8985	0.9034	0.9105
	MCC	0.8997	0.9561	0.9484	0.8992	0.9132	0.9130	0.9216
	OA	0.9495	0.9781	0.9758	0.9529	0.9581	0.9574	0.9620
SC-RoadDeepNet (α=0.7)	F1 score	0.9398	0.9687	0.9611	0.9283	0.9475	0.9491	0.9491
	IoU	0.8864	0.9392	0.9251	0.8661	0.9002	0.9031	0.9034
	MCC	0.8934	0.9390	0.9373	0.8916	0.9146	0.9129	0.9148
	OA	0.9458	0.9695	0.9705	0.9498	0.9591	0.9575	0.9587
SC-RoadDeepNet (α=0.9)	F1 score	0.9367	0.9711	0.9495	0.9311	0.9457	0.9441	0.9464
	IoU	0.8809	0.9438	0.9039	0.8710	0.8970	0.8941	0.8985
	MCC	0.8876	0.9439	0.9201	0.8956	0.9127	0.9052	0.9109
	OA	0.9441	0.9720	0.9625	0.9521	0.9589	0.9541	0.9573

TABLE 2. QUANTITATIVE EXPERIMENTAL OUTCOMES YIELDED BY THE RRCNN APPROACH FOR ROAD EXTRACTION WITHOUT BL AND CP_CLDICE TECHNIQUES

		Image 1	Image 2	Image 3	Image 4	Image 5	Image 6	Average
RRCNN	F1 score	0.9350	0.9424	0.9513	0.9140	0.9386	0.9301	0.9352
	IoU	0.8779	0.8909	0.9071	0.8415	0.8842	0.8693	0.8785
	MCC	0.8853	0.8876	0.9227	0.8694	0.9000	0.8807	0.8910
	OA	0.9407	0.9438	0.9637	0.9402	0.9522	0.9421	0.9471

V. DISCUSSION

In this section, we evaluated the performance of the proposed framework by analyzing the ablation study and testing the model on another road dataset.

A. Ablation study

To assess the efficiency of the proposed shape and connectivity-preserving the ability of the SC-RoadDeepNet model to improve road discontinuity and road shape segmentation, we conducted an ablation study in this work. In this case, we applied the proposed RRCNN model with the primary binary cross-entropy loss function and without BL and CP_clDice techniques to see the influence of these methods on fine-tuning road shape and preserving road connectivity. We obtained the quantitative and visualization findings by the model in road segmentation from the Google Earth dataset. Table 2 contains the quantitative results, whereas Figure 10 depicts the visualization results. After adjusting various variables and removing those crucial techniques, the accuracy of the IoU in the proposed RRCNN model was reduced to 87.85%, as shown in Table 2. Furthermore, as shown in Figure 14, the suggested approach introduced spurs and generated more FPs and FNs in homogeneous areas, thus reducing the smoothness and connectedness of the road segmentation network significantly. Therefore, BL and CP_clDice showed a significant role in preserving road shape and connectivity and producing high-quality road segmentation maps.

B. DeepGlobe and Massachusetts road datasets

Furthermore, we applied our proposed SC-RoadDeepNet model on more road datasets called DeepGlobe [42] and Massachusetts [43] to show the model's efficiency in road segmentation from various types of remote sensing imagery. The DeepGlobe dataset was captured in India, Indonesia, and Thailand, which contained 8570 images with 50 cm per pixel spatial resolution and covered 2220 km^2. Each image was 1024×1024 pixels in size. The training and testing datasets consisted of 1006 and 26 images in this study, respectively. The Massachusetts dataset that we used contained 1032 training and 32 testing images with a size of 768×768 and spatial resolution of 0.5 m. We obtained quantitative and visualization outcomes yielded by the presented RRCNN, RRCNN+BL, RRCNN+CP_clDice, and SC-RoadDeepNet models for road segmentation from the DeepGlobe and Massachusetts datasets, which are demonstrated in Table 3 and Figure 11 for the DeepGlobe dataset and Table 4 and Figure 12 for the Massachusetts dataset, respectively. Table 3 shows that the proposed RRCNN model did not benefit from BL and CP_clDice techniques achieved the lowest F1 score with 91.49% for DeepGlobe and 87.19% for Massachusetts.

In contrast, the proposed RRCNN+BL, RRCNN+CP_clDice, and SC-RoadDeepNet could improve the results of DeepGlobe to 92.08%, 92.30%, and 92.78% for F1 score and the results of Massachusetts to 87.95%, 88.47%, and 89.33%, respectively. According to the visualization results (Figures 11 and 12), the proposed RRCNN model failed to segment roads in the complex areas, where the road was covered by shadows and trees and brought in more FPs, FNs, and discontinuity.

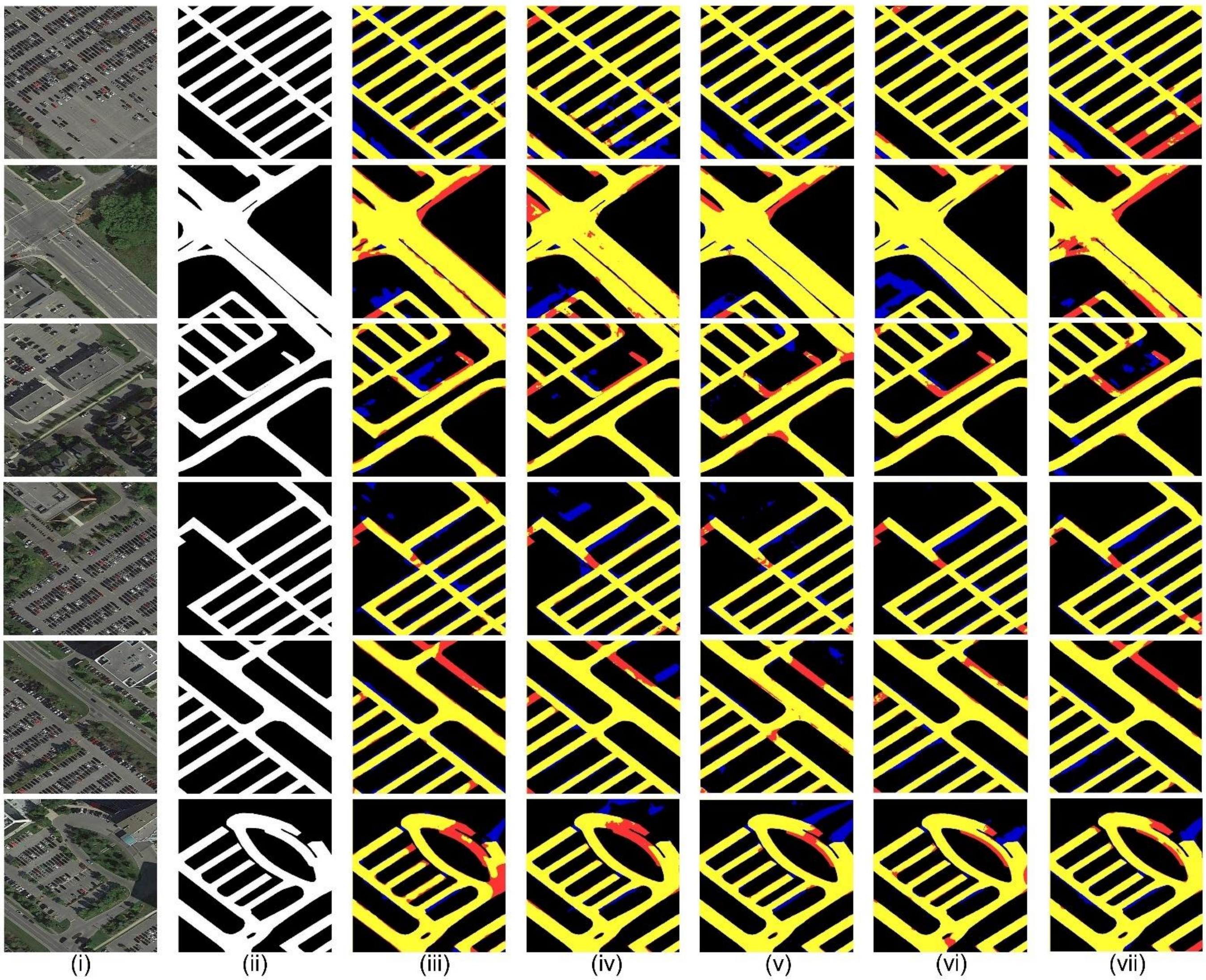
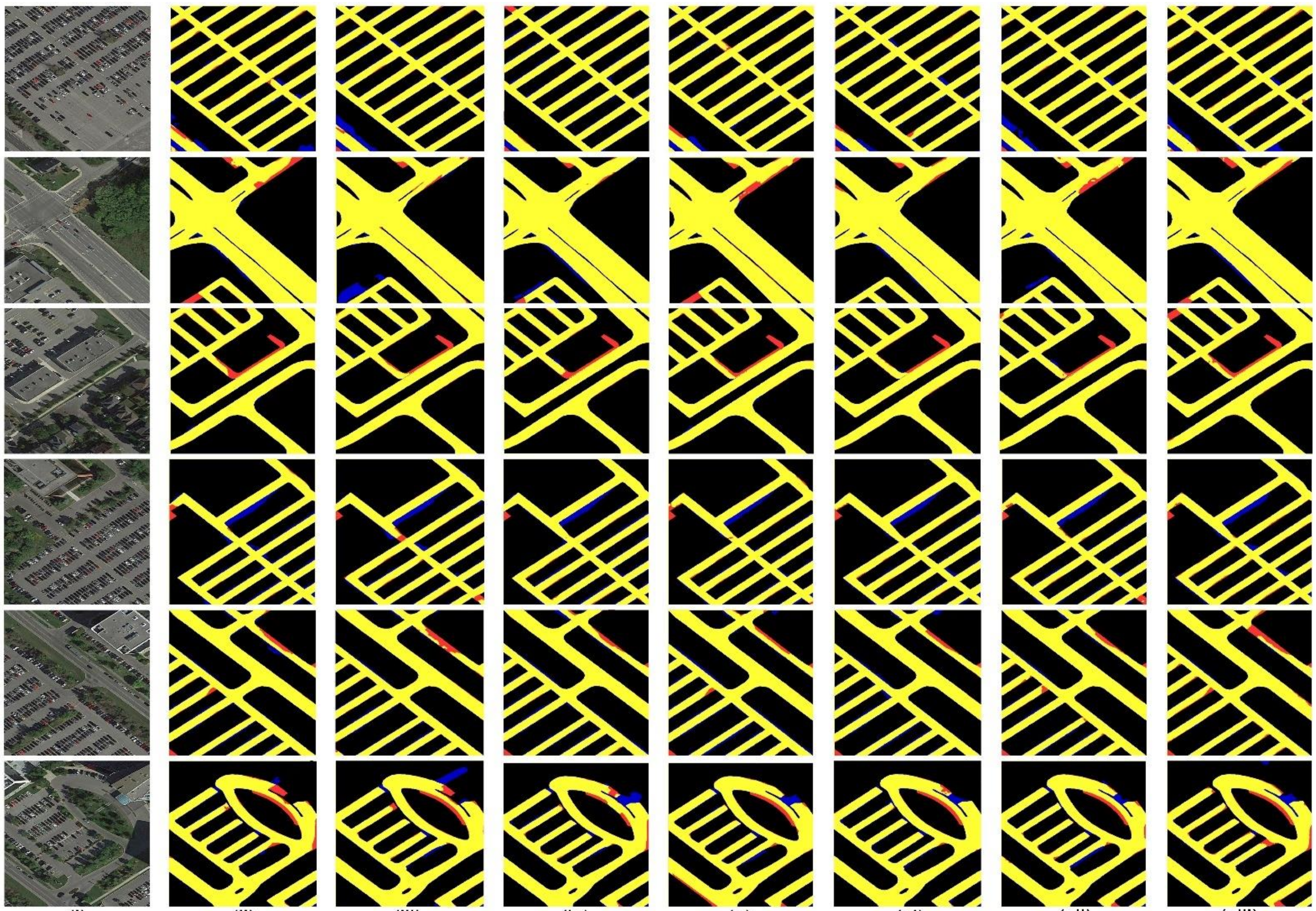


Fig. 8. Road qualitative results were compared visually using various comparing models: (i) original RGB Google Earth images, (ii) reference images, (iii) LinkNet results, (iv) ResUNet results, (v) UNet results, (vi) VNet results, and (vii) DeeplabV3+ results. TPs, FPs, and FNs are represented by yellow, blue, and red, respectively.

TABLE 3. QUANTITATIVE EXPERIMENTAL OUTCOMES YIELDED BY THE RRCNN, RRCNN+BL, RRCNN+CP_CLDICE, AND SC-ROADDEEPNET APPROACHES FOR ROAD EXTRACTION FROM THE DEEPGLOBE ROAD DATASET

		Image 1	Image 2	Image 3	Image 4	Image 5	Average
RRCNN	F1 score	0.9207	0.9250	0.8927	0.8918	0.9444	0.9149
	IoU	0.8529	0.8604	0.8061	0.8047	0.8947	0.8438
	MCC	0.9078	0.9157	0.8632	0.8763	0.9329	0.8992
	OA	0.9774	0.9835	0.9534	0.9722	0.9808	0.9735
RRCNN+BL	F1 score	0.9327	0.9296	0.8976	0.8973	0.9469	0.9208
	IoU	0.8738	0.8684	0.8141	0.8137	0.8990	0.8538
	MCC	0.9215	0.9209	0.8707	0.8823	0.9360	0.9063
	OA	0.9807	0.9846	0.9563	0.9734	0.9817	0.9753
RRCNN+CP_clDice	F1 score	0.9394	0.9304	0.8983	0.8988	0.9481	0.9230

	IoU	0.8858	0.8698	0.8154	0.8161	0.9013	0.8577
	MCC	0.9294	0.9217	0.8726	0.8834	0.9380	0.9090
	OA	0.9828	0.9846	0.9570	0.9733	0.9824	0.9760
SC-RoadDeepNet	F1 score	0.9416	0.9349	0.9062	0.9065	0.9499	0.9278
	IoU	0.8896	0.8777	0.8285	0.8289	0.9046	0.8659
	MCC	0.9319	0.9271	0.8805	0.8924	0.9394	0.9143
	OA	0.9834	0.9859	0.9593	0.9756	0.9826	0.9774



(i) (ii) (iii) (iv) (v) (vi) (vii) (viii)

Fig. 9. Road qualitative results were compared visually using proposed models: (i) original RGB Google Earth images, (ii) RRCNN+BL results, (iii) RRCNN+CP_clDice results, (iv) SC-RoadDeepNet results (α=0.1), (v) SC-RoadDeepNet results (α=0.3), (vi) SC-RoadDeepNet results (α=0.5), (vii) SC-RoadDeepNet results (α=0.7), and (viii) SC-RoadDeepNet results, (α=0.9). The TPs, FPs, and FNs are represented by yellow, blue, and red, respectively.

In contrast, the presented SC-RoadDeepNet that benefited from BL and CP_clDice could obtain the segmentation map with fewer FPs and FNs and showed higher extraction accuracy on the boundary and road connectivity than others. In summary, the proposed method could improve road extraction by tackling occlusion-related interruptions. It could solve discontinuity in road extraction results and produce high-resolution results compared with the other methods. We also calculated the runtime of the presented method on each dataset, which took 117 s, 388 s, and 226 s per epoch for the training process for the Ottawa, DeepGlobe, and Massachusetts datasets, respectively. The model was trained for 100 epochs. Therefore, it took 195 minutes for the Ottawa dataset, 646.66 minutes for the DeepGlobe dataset, and 376.66 minutes for the Massachusetts dataset. As the size of images and datasets increased, the training time also increased. Overall, the suggested method did not need a huge training dataset or a lot of computational effort, yet it still outperformed previous models in terms of statistical outcomes.

TABLE 4. QUANTITATIVE EXPERIMENTAL OUTCOMES YIELDED BY THE RRCNN, RRCNN+BL, RRCNN+CP_CLDICE, AND SC-ROADDEEPNET APPROACHES FOR ROAD EXTRACTION FROM THE MASSACHUSETTS ROAD DATASET

		Image 1	Image 2	Image 3	Image 4	Image 5	Average
RRCNN	F1 score	0.8827	0.8591	0.8785	0.8663	0.8730	0.8719
	IoU	0.8099	0.7729	0.8032	0.7841	0.7946	0.7929
	MCC	0.8586	0.8320	0.8614	0.8490	0.8543	0.8511
	OA	0.9552	0.9534	0.9680	0.9677	0.9642	0.9617
RRCNN+BL	F1 score	0.8964	0.8711	0.8866	0.8700	0.8733	0.8795
	IoU	0.8321	0.7915	0.8162	0.7898	0.7950	0.8049
	MCC	0.8738	0.8477	0.8704	0.8529	0.8538	0.8597
	OA	0.9627	0.9599	0.9716	0.9698	0.9663	0.9661
RRCNN+CP_clDice	F1 score	0.8985	0.8743	0.8898	0.8820	0.8790	0.8847
	IoU	0.8357	0.7966	0.8215	0.8088	0.8040	0.8133
	MCC	0.8765	0.8518	0.8743	0.8665	0.8604	0.8659
	OA	0.9629	0.9611	0.9726	0.9726	0.9678	0.9674
SC-RoadDeepNet	F1 score	0.9037	0.8808	0.9039	0.8899	0.8881	0.8933
	IoU	0.8443	0.8070	0.8446	0.8216	0.8187	0.8272
	MCC	0.8828	0.8581	0.8902	0.8762	0.871	0.8757
	OA	0.9655	0.9617	0.976	0.9752	0.9695	0.9696

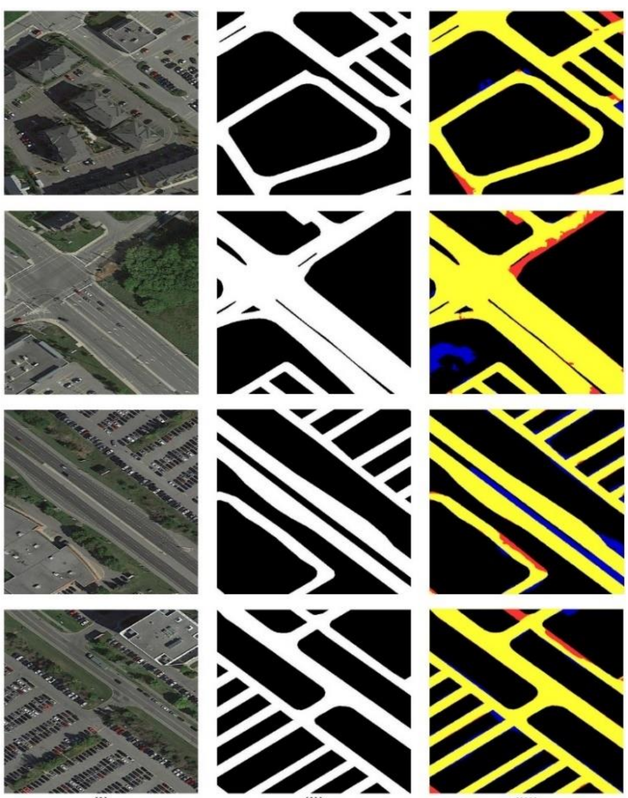


Fig. 10. Road qualitative results were compared visually using the proposed RRCNN model: (i) original RGB Google Earth images, (ii) reference images, (iii) RRCNN results. TPs, FPs, and FNs are represented by yellow, blue, and red, respectively.

VI. CONCLUSION

This study introduced SC-RoadDeepNet, a new method for extracting roads from remote sensing imagery based on a shape and connectivity-preserving road segmentation deep learning model. The proposed model consisted of a state-of-the-art deep learning model called the RRCNN model, BL, and CP_clDice techniques. The RRCNN model included convolutional encoder-decoder units similar to the primary UNet model. However, in the encoder-decoder arms, RRCLs were used instead of standard forward convolutional layers. RRCLs aided in the development of a more effective deeper structure. Furthermore, the suggested model's RRCL units provided an effective feature accumulation mechanism. Concerning distinct time-steps, feature accumulation guaranteed stronger and better feature representation. As a result, it aided in the extraction of low-level features that are critical for segmentation tasks. We also used BL to punish boundary misclassification and fine-tune the road form as a result. We provided CP_clDice for maintaining road connectivity and obtaining correct segmentations. The suggested framework was tested on high-resolution remote sensing datasets, and the findings demonstrated its usefulness and feasibility in increasing the performance of road semantic segmentation. Qualitative comparisons were compared with several comparative semantic segmentation algorithms. The presented model outperformed the other models, thus preserving shape and road connectivity and achieving high-resolution segmentation maps according to the results of the experiments. Compared with the aforementioned semantic segmentation methods, the suggested method could also improve the complete assessment metrics, such as the IoU and F1 score.

FUNDING

This research was funded by the Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), Faculty of Engineering and Information Technology, the University of Technology Sydney, Australia. This research was also supported by the Researchers Supporting Project number RSP-2021/14, King Saud University, Riyadh, Saudi Arabia.

AUTHOR CONTRIBUTIONS

Conceptualization: A.A. and B.P.; methodology and formal analysis: A.A.; data curation, A.A.; writing—original draft preparation: A.A.; Validation: B.P.; Visualization: B.P.; Resources: B.P., writing—review and editing, B.P., A.Almri; Supervision: B.P.; and funding: B.P., A. Alamri.

DATA AVAILABILITY

The link to download the Google Earth, Massachusetts, and DeepGlobe road datasets can be found at https://github.com/yhlleo/RoadNet, https://www.cs.toronto.edu/~vmnih/data/, and https://www.kaggle.com/balraj98/deepglobe-road-extraction-dataset.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

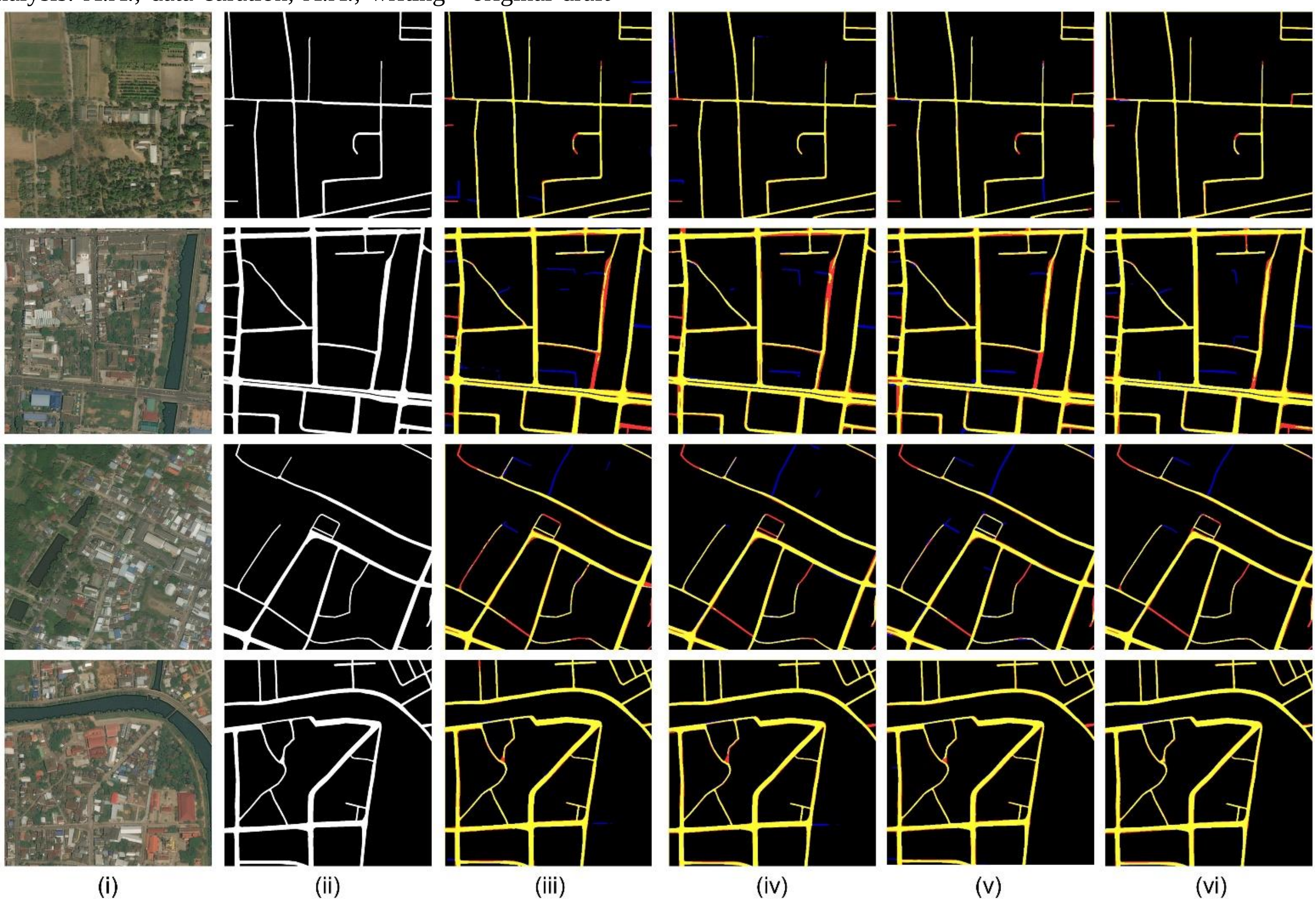


Fig. 11. Road qualitative results achieved by the models from the DeepGlobe road dataset: (i) original RGB images, (ii) reference images, (iii) RRCNN results, (iv) RRCNN+BL results, (v) RRCNN+CP_clDice results, and (vi) SC-RoadDeepNet results. TPs, FPs, and FNs are represented by yellow, blue, and red, respectively.

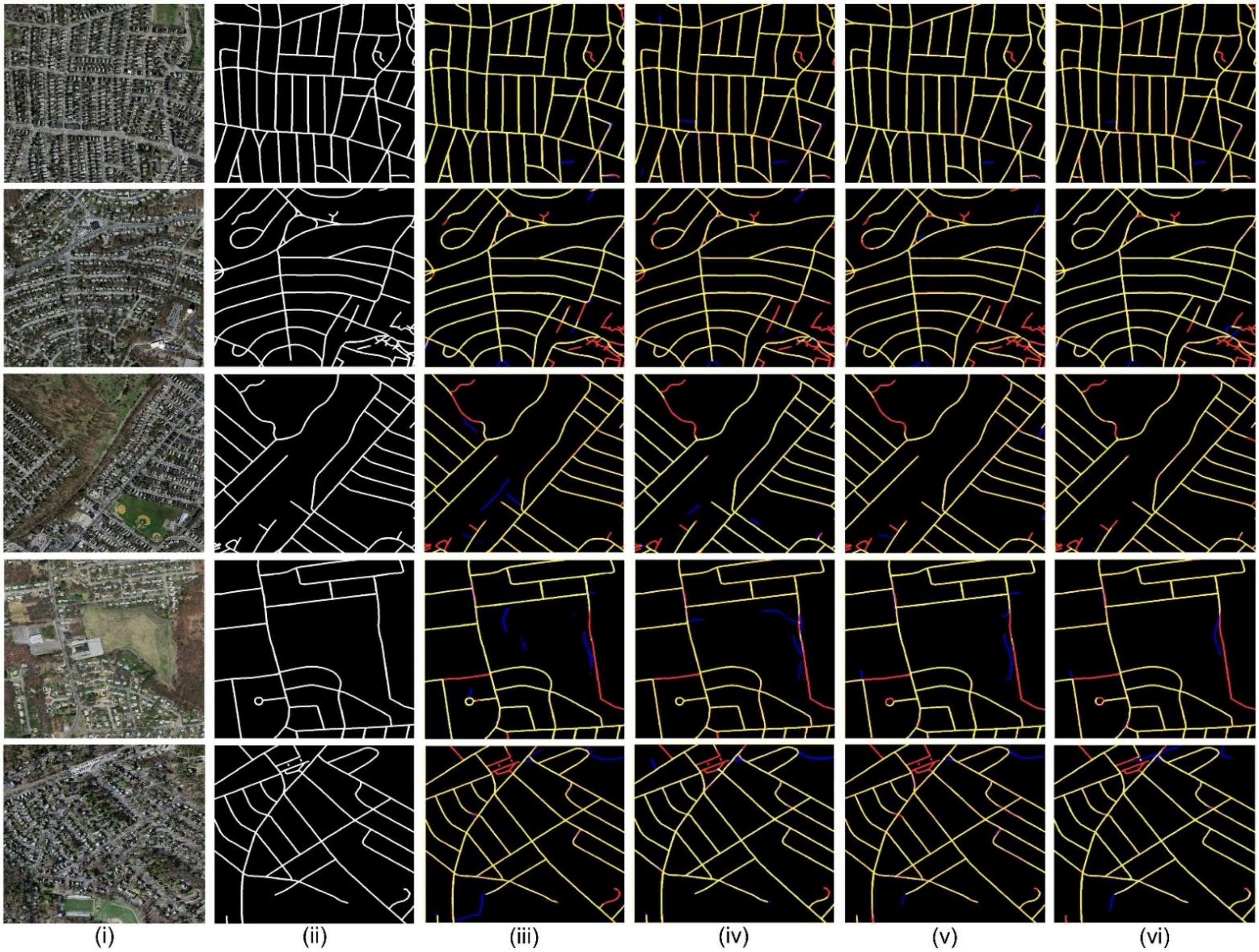


Fig. 12. Road qualitative results achieved by the models from the Massachusetts road dataset: (i) original RGB images, (ii) reference images, (iii) RRCNN results, (iv) RRCNN+BL results, (v) RRCNN+CP_clDice results, and (vi) SC-RoadDeepNet results. TPs, FPs, and FNs are represented by yellow, blue, and red, respectively.

REFERENCE

[1] Y. Ma *et al.*, "Remote sensing big data computing: Challenges and opportunities," *Future Generation Computer Systems*, vol. 51, pp. 47-60, 2015.

[2] P. Liu, L. Di, Q. Du, and L. Wang, "Remote Sensing Big Data: Theory, Methods and Applications," *Remote Sensing*, vol. 10, no. 5, p. 711, 2018. [Online]. Available: https://www.mdpi.com/2072-4292/10/5/711.

[3] F. Casu, M. Manunta, P. Agram, and R. Crippen, "Big Remotely Sensed Data: tools, applications and experiences," *Remote Sensing of Environment*, vol. 202, no. 1, pp. 1-2, 2017.

[4] A. Abdollahi, B. Pradhan, and A. Alamri, "RoadVecNet: a new approach for simultaneous road network segmentation and vectorization from aerial and google earth imagery in a complex urban set-up," *GIScience & Remote Sensing*, pp. 1-24, 2021, doi: 10.1080/15481603.2021.1972713.

[5] J. Wang, J. Song, M. Chen, and Z. Yang, "Road network extraction: A neural-dynamic framework based on deep learning and a finite state machine," *International Journal of Remote Sensing*, vol. 36, no. 12, pp. 3144-3169, 2015.

[6] M. O. Sghaier and R. Lepage, "Road extraction from very high resolution remote sensing optical images based on texture analysis and beamlet transform," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 5, pp. 1946-1958, 2015.

[7] Z. Miao, B. Wang, W. Shi, and H. Zhang, "A semi-automatic method for road centerline extraction from VHR images," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 11, pp. 1856-1860, 2014.

[8] C. Unsalan and B. Sirmacek, "Road network detection using probabilistic and graph theoretical methods," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 11, pp. 4441-4453, 2012, doi: https://doi.org/10.1109/TGRS.2012.2190078.

[9] H. R. R. Bakhtiari, A. Abdollahi, and H. Rezaeian, "Semi automatic road extraction from digital images," *The Egyptian Journal of Remote Sensing and Space Science*, vol. 20, no. 1, pp. 117-123, 2017/06/01/ 2017, doi: https://doi.org/10.1016/j.ejrs.2017.03.001.

[10] R. Alshehhi and P. R. Marpu, "Hierarchical graph-based segmentation for extracting road networks from high-resolution satellite images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 126, pp. 245-260, 2017/04/01/ 2017, doi: https://doi.org/10.1016/j.isprsjprs.2017.02.008.

[11] S. Das, T. Mirnalinee, and K. Varghese, "Use of salient features for the design of a multistage framework to extract roads from high-resolution multispectral satellite images," *IEEE transactions on Geoscience Remote sensing*, vol. 49, no. 10, pp. 3906-3931, 2011.

[12] M. Song and D. Civco, "Road extraction using SVM and image segmentation," *Photogrammetric Engineering & Remote Sensing*, vol. 70, no. 12, pp. 1365-1371, 2004.

[13] S. Wang, X. Mu, D. Yang, H. He, and P. Zhao, "Road extraction from remote sensing images using the inner convolution integrated encoder-decoder network and directional conditional random fields," *Remote Sensing*, vol. 13, no. 3, p. 465, 2021.

[14] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436-444, 2015.

[15] D. Hong *et al.*, "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 5, pp. 4340-4354, 2020.

[16] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, 2020.

[17] D. Hong *et al.*, "Endmember-Guided unmixing network (EGU-Net): A general deep learning framework for self-supervised hyperspectral unmixing," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[18] R. Hang, Z. Li, P. Ghamisi, D. Hong, G. Xia, and Q. Liu, "Classification of hyperspectral and LiDAR data using coupled CNNs," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 7, pp. 4939-4950, 2020.

[19] A. Abdollahi, B. Pradhan, N. Shukla, S. Chakraborty, and A. Alamri, "Multi-Object segmentation in complex urban scenes from high-resolution remote sensing data," *Remote Sensing*, vol. 13, no. 18, p. 3710, 2021.

[20] V. Mnih and G. E. Hinton, "Learning to detect roads in high-resolution aerial images," Berlin, Heidelberg, 210-223, 2010. https://doi.org/10.1007/978-3-642-15567-3_16.

[21] M. Rezaee and Y. Zhang, "Road detection using deep neural network in high spatial resolution images," in *2017 Joint Urban Remote Sensing Event (JURSE)*, 2017: IEEE, pp. 1-4.

[22] A. Abdollahi, B. Pradhan, G. Sharma, K. N. A. Maulud, and A. Alamri, "Improving road semantic segmentation using generative adversarial network," *IEEE Access*, 64381 - 64392, 2021.

[23] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 801-818.

[24] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234-241. https://doi.org/10.1007/978-3-319-24574-4_28.

[25] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis Machine Intelligence*, vol. 39, no. 12, pp. 2481-2495, 2017.

[26] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, vol. 39, no. 4, pp. 3431-3440, doi: https://doi.org/10.1109/TPAMI.2016.2572683.

[27] Y. Li, B. Peng, L. He, K. Fan, Z. Li, and L. Tong, "Road extraction from unmanned aerial vehicle remote sensing images based on improved neural networks," *Sensors*, vol. 19, no. 19, p. 4115, 2019.

[28] Z. L. Zhang, Qingjie; Wang, Yunhong, "Road extraction by deep residual U-Net," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 5, pp. 749-753, 2018, doi: 10.1109/LGRS.2018.2802944.

[29] Z. Zhang and Y. Wang, "JointNet: A common neural network for road and building extraction," *Remote Sensing*, vol. 11, no. 6, p. 696, 2019.

[30] Z. Zhong, J. Li, W. Cui, and H. Jiang, "Fully convolutional networks for building and road

extraction: preliminary results," in *Geoscience and Remote Sensing Symposium (IGARSS), IEEE International, 1591-1594,* 2016.

[31] H. He, D. Yang, S. Wang, S. Wang, and Y. Li, "Road extraction by using atrous spatial pyramid pooling integrated encoder-decoder network and structural similarity loss," *Remote Sensing*, vol. 11, no. 9, p. 1015, 2019.

[32] A. Abdollahi, B. Pradhan, and A. Alamri, "VNet: An end-to-end fully convolutional neural network for road extraction from high-resolution remote sensing data," *IEEE Access*, vol. 8, pp. 179424 - 179436, 2020.

[33] A. Mosinska, P. Marquez-Neila, M. Koziński, and P. Fua, "Beyond the pixel-wise loss for topology-aware delineation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3136-3145.

[34] Y. Liu, J. Yao, X. Lu, M. Xia, X. Wang, and Y. Liu, "Roadnet: Learning to comprehensively analyze road networks in complex urban scenes from high-resolution remotely sensed images," *IEEE Transactions on Geoscience Remote Sensing*, vol. 57, no. 4, pp. 2043-2056, 2018.

[35] M. Liang and X. Hu, "Recurrent convolutional neural network for object recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3367-3375.

[36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770-778.

[37] F. I. Diakogiannis, F. Waldner, P. Caccetta, C. J. I. J. o. P. Wu, and R. Sensing, "Resunet: a deep learning framework for semantic segmentation of remotely sensed data," vol. 162, pp. 94-114, 2020.

[38] K. Palágyi, "A 3-subiteration 3D thinning algorithm for extracting medial surfaces," *Pattern Recognition Letters*, vol. 23, no. 6, pp. 663-675, 2002.

[39] F. Y. Shih and C. C. Pu, "A skeletonization algorithm by maxima tracking on Euclidean distance transform," *Pattern Recognition,* vol. 28, no. 3, pp. 331-341, 1995.

[40] N. Ghasemkhani, S. S. Vayghan, A. Abdollahi, B. Pradhan, and A. Alamri, "Urban development modeling using integrated fuzzy systems, ordered weighted averaging (OWA), and geospatial techniques," *Sustainability,* vol. 12, no. 3, p. 809, 2020.

[41] A. Chaurasia and E. Culurciello, "Linknet: Exploiting encoder representations for efficient semantic segmentation," in *2017 IEEE Visual Communications and Image Processing (VCIP),* 2017, pp. 1-4.

[42] I. Demir *et al.*, "Deepglobe: A challenge to parse the earth through satellite images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 172-181.

[43] V. Mnih, *Machine learning for aerial image labeling, Ph.D. dissertation, Dept. Comput. Sci., Univ. Toronto, Toronto, ON, Canada*. Citeseer, 2013.

ABOLFAZL ABDOLLAHI received a B.Sc degree from Ferdowsi University of Mashhad, Iran and an M.Sc degree in GIS and Remote Sensing from Kharazmi University of Tehran, Iran. He is currently a PhD student with the Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), University of Technology Sydney (UTS). His research interests contain the application of advance machine learning approaches and deep learning-based networks for remote sensing image classification, image segmentation, feature extraction, and GIS maps database updating. He was rewarded the International Research Scholarship and UTS Presidents' Scholarship for the current course in 2018. He has published numerous peer-reviewed papers on the application of machine learning approaches.

BISWAJEET PRADHAN (M'12, SM'16) received a Habilitation degree in remote sensing from the Dresden University of Technology, Germany in 2011. He is currently the Director of the Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), Faculty of Engineering and IT. He is also a Distinguished Professor with the University of Technology Sydney. He is also an internationally established scientist in the fields of geospatial information systems (GIS), remote sensing and image processing, complex modeling/geo-computing, machine learning and soft-computing applications, natural hazards, and environmental modeling. In 2015–2021, he served as the Ambassador Scientist for the Alexander Humboldt Foundation, Germany. Out of his more than 650 articles, over 500 have been published in science citation index (SCI/SCIE) technical journals. He has authored eight books and 13 book chapters. He was also a recipient of the Alexander von Humboldt Fellowship from Germany. He received 55 awards in recognition of his excellence in teaching, service, and research, since 2006. From 2016 to 2020, he was listed as the World's Most Highly Cited Researcher by Clarivate Analytics Report and as one of the world's most influential minds. In 2018–2020, he was awarded as the World Class Professor by the Ministry of Research, Technology and Higher Education, Indonesia. He is also an Associate Editor and an Editorial Member of more than eight ISI journals. He has widely travelled abroad, visiting more than 52 countries to present his research findings.

ABULLAH ALAMRI, M.S., is a professor of earthquake seismology and is the Director of Seismic Studies Center at King Saud University (KSU). He is the President of the Saudi Society of Geosciences and editor-in-chief of the Arabian Journal of Geosciences (AJGS). He holds a B.S. in geology (1981) from King Saud University, M.Sc. (1985) in applied geophysics from University of South Florida, Tampa and Ph.D (1990) in earthquake seismology from University of Minnesota, USA. He is a member of Seismological Society of America, American Geophysical Union, European Ass. for Environmental and Eng. Geophysics, Earthquakes Mitigation in the Eastern Mediterranean Region, National Comm. for Assessment and Mitigation of Earthquake Hazards in

Saudi Arabia, and Mitigation of Natural Hazards Com at Civil Defense. His research interests are in the area of crustal structures and seismic micro zoning of the Arabian Peninsula. His recent projects also involve applications of EM and MT in deep groundwater exploration of Empty Quarter and geothermal prospecting of volcanic Harrats in the Arabian shield. He has published more than 150 research papers, achieved more than 45 research projects as well as authored several books and technical reports. He is a principal and co-investigator in several national and international projects (KSU, KACST, NPST, IRIS, CTBTO, US Air Force, NSF, UCSD, LLNL, OSU, PSU, and Max Planck). He has also chaired and co-chaired several SSG, GSF, and RELEMR workshops and forums in the Middle East. He obtained several worldwide prizes and awards for his scientific excellence and innovation.